Symbols and Computation
A Critique of the Computational Theory of Mind

STEVEN HORST
Wesleyan University

Abstract. Over the past several decades, the philosophical community has witnessed the emergence of an important new paradigm for understanding the mind.[1] The paradigm is that of machine computation, and its influence has been felt not only in philosophy, but also in all of the empirical disciplines devoted to the study of cognition. Of the several strategies for applying the resources provided by computer and cognitive science to the philosophy of mind, the one that has gained the most attention from philosophers has been the 'Computational Theory of Mind' (CTM). CTM was first articulated by Hilary Putnam (1960, 1961), but finds perhaps its most consistent and enduring advocate in Jerry Fodor (1975, 1980, 1981, 1987, 1990, 1994). It is this theory, and not any broader interpretations of what it would be for the mind to be a computer, that I wish to address in this paper. What I shall argue here is that the notion of 'symbolic representation' employed by CTM is fundamentally unsuited to providing an explanation of the intentionality of mental states (a major goal of CTM), and that this result undercuts a second major goal of CTM, sometimes refered to as the 'vindication of intentional psychology.' This line of argument is related to the discussions of 'derived intentionality' by Searle (1980, 1983, 1984) and Sayre (1986, 1987). But whereas those discussions seem to be concerned with the *causal dependence* of familiar sorts of symbolic representation upon meaning-bestowing acts, my claim is rather that there is not one but several notions of 'meaning' to be had, and that the notions that are applicable to symbols are *conceptually dependent* upon the notion that is applicable to mental states in the fashion that Aristotle refered to as *paronymy*. That is, an analysis of the notions of 'meaning' applicable to symbols reveals that they contain presuppositions about meaningful mental states, much as Aristotle's analysis of the sense of 'healthy' that is applied to foods reveals that it means 'conducive to having a *healthy body*,' and hence any attempt to explain 'mental semantics' in terms of the semantics of symbols is doomed to circularity and regress. I shall argue, however, that this does not have the consequence that computationalism is bankrupt as a paradigm for cognitive science, as it is possible to reconstruct CTM in a fashion that avoids these difficulties and makes it a viable research framework for psychology, albeit at the cost of losing its claims to explain intentionality and to vindicate intentional psychology. I have argued elsewhere (Horst, 1996) that local special sciences such as psychology do not require vindication in the form of demonstrating their reducibility to more fundamental theories, and hence failure to make good on these philosophical promises need not compromise the broad range of work in empirical cognitive science motivated by the computer paradigm in ways that do not depend on these problematic treatments of symbols.

1. The Computational Theory of Mind

CTM is made up of two components: a *representational* account of the nature of intentional states, and a *computational* account of cognitive processes. CTM claims, first, that intentional states such as token beliefs and desires are *relational* states involving a cognizer and a *mental representation*. This *mental* representation is a form of *symbolic* representation: mental representations are physically-instantiated

Minds and Machines **9:** 347–381, 1999.

symbol tokens having both syntactic and semantic properties. (cf. Fodor 1981, p. 26) Fodor indeed speaks of the system of mental representations as being quite literally a 'language of thought.'

Viewing mental representation on the model of symbolic representation in a language is thought to have three principal merits. First, it provides the basis for an account of the intentionality and semantic properties of mental states. Symbols and mental states, claims Fodor, are the only things that have semantic properties; and so it might be plausible to suppose that the long-lived philosophical intuition that thought is representational can be cashed out in symbolic terms, such that the semantic properties of mental states are 'inherited' from those of the mental representations they contain. (1987, p. xi) A second advantage of viewing mental representation on the model of language is that this provides the resources for endowing thought with the same generative and creative aspects possessed by languages. (Fodor, 1987, appendix.) Not all symbol structures in a language are semantic atoms—most are derived by the application of a finite set of compositional rules to a finite number of lexical primitives. Yet such rules allow for the generation of an infinite variety of senses out a finite number of rules and lexical primitives. Indeed, the *only* known way of providing this kind of generative capacity is through having a language containing items that are lexical primitives plus a set of syntactically-based compositional rules. So the best working hypothesis for endowing *thought* with this kind of capacity would seem to be to suppose that thought involves symbolic representations, some of which are semantically primitive, and to suppose that these mental representations have syntactic properties that allow for the application of compositional rules that can generate semantically complex symbol-structures as well.

The third advantage of CTM—and more specifically, of the language of thought hypothesis—is that it provides a computational account of mental processes. What representational accounts of thought traditionally have lacked is a way of accounting for mental *processes* such as reasoning to a conclusion, since representations are semantically-typed, and there is no obvious way of building a causal nomological theory based around semantically-typed items. Here, however, two developments over the last century have provided the basis for a dramatic paradigm shift. Discussions of *formalization* in late nineteenth-century mathematics showed that, for significant (albeit limited) domains, it is possible to devise a set of representations and derivational rules such that the rules allow for all and only the deductions one would want on semantic grounds, while the rules themselves are sensitive only to the syntactic form of the expressions, not to their semantic values. The second development, machine computation, has shown, roughly, that any finite formalizable procedure can be implemented by a digital machine. In somewhat colloquial terms, formalization shows how to link up semantics with syntax, and the computer paradigm shows how to link up the syntactic properties of symbols with causal powers. And this provides inspiration for an approach to mental processes: if the locus of the semantic properties of mental states is the representations they contain, it might be the case that (1) all semantic distinctions between mental representations are mirrored by syntactic distinctions, and (2) it is the syntactic properties of the symbols that are relevant to their causal contribution to mental processes. If this is the case, the semantic relationships that seem crucial to a chain of reasoning can have a causal grounding, because semantics is coordinated with syntax, and syntax determines causal powers. (Fodor 1987, pp. 19–20)

CTM thus seeks to supply two things that are of signal interest to the philosophy of mind: (1) a representational account of the intentionality and semantic properties of mental states, and (2) an explanation of mental processes that can provide a 'vindication' of psychological explanations of mental processes cast in a mentalistic or intentional vocabulary—that is, a vindication of intentional psychology. It should be clear that the second project is dependent upon the first. In order to link the semantic properties of mental states to causal laws by way of computation, it is necessary to suppose (i) that there are mental represenions whose semantic properties are coordinated with their syntax, and (ii) that it is these mental representations that are the locus of the semantic properties of intentional states. A *purely* computational-syntactic account of psychological processes *without* a representational component, such as that of Stich (1983) might have independent interest as well, but in order to be used to justify explanation in terms of beliefs and desires, it must be coupled with an account of representational semantics that can ground the semantics and intentionality of mental states. I shall thus confine myself to a discussion of CTM's representational account of mental states. If the considerations I shall raise are decisive against this account, the further goal of vindicating intentional psychology by combining representation with computation is a non-starter.

2. Derived Intentionality

One important line of criticism that has been developed against CTM is based upon an intuition that there is something fundamentally flawed in the strategy of explaining the meaningfulness of mental states by appealing to meaningful symbolic representations. The problem, simply stated, is that CTM has the relationship between mental meaning and symbolic meaning precisely reversed. For when one is required to give an account of the meaningfulness of the symbols employed in a language, one cannot do so except by appeal to the conventions of communities of language-users, the intentions of speakers and writers, and the interpretive acts of readers and listeners. But if the meaningfulness of symbols in a language can only be explained in a fashion that invokes meaningful mental states, then CTM's strategy for explaining the meaningfulness of mental states by appeal to meaningful symbols is doomed to circularity and regress. It is doomed to *circularity* because the meaningfulness of mental states is explained by reference to the meanings of symbols while the meaningfulness of symbols must in turn be explained by reference to conventions and intentions of symbol-users (and hence ultimately to other meaningful mental states). It is consigned to *regress* because the explanation of the meaning of each mental state M will eventually have to be cashed out in terms of other mental states–namely, those states M^* that are used to explain the meaning of the mental representation R that is supposed to account for the meaning of M.

One view of this issue of the relationship between the meaningfulness of mental states and that of symbols is often discussed under the heading of 'derived intentionality.' (Searle 1981, 1984; Sayre 1986, 1987). According to this view, the meaningfulness and intentionality of symbols is *derived* from that of the mental states of which they are expressions. Proponents of this view characteristically believe that the meaningfulness and intentionality of mental states is not similarly derived, but *original* or *intrinsic*.

There are, however, two very different lines of criticism to be had here, each of which is suggested by the expression 'derived intentionality'. Each raises serious difficulties for CTM, but the nature and force of these difficulties differs significantly. The first and more familiar line of criticism, which I shall call the 'Causal Derivation Objection,' locates the problem for CTM in the fact that the the intentionality of symbols such as inscriptions and illocutionary acts is causally dependent upon pre-existing intentional states and upon meaning-bestowing acts. The second line of criticism, the 'Conceptual Dependence Objection,' proceeds differently. According to this view, the very *notion* of 'symbolic meaning' is *conceptually dependent* upon a distinct notion of 'mental meaning,' in much the fashion that the usage of the word 'healthy' that is applied to foods is conceptually dependent upon the usage of 'healthy' that is applied to living bodies. These are two significantly different lines of argument. The first assumes that there is nothing intrinsically conventional about *meaning* (or other semantic properties) *per se*, and that semantic properties may be predicated univocally of mental states, discursive symbols and mental representations. The second view, by contrast, holds that there are several things that go by the name of 'meaning', and that those which may be attributed to symbols are conventional to the core, while those that may be attributed to mental states are not. In the following sections I shall develop each of these objections and attempt to assess how each succeeds as a criticism of CTM.

3. Causally Derived Intentionality

The first objection to CTM is familiar from several writers who have made use of the notion of 'derived intentionality.' John Searle's (1980, 1984) arguments against CTM are perhaps the best known examples of this line of criticism. According to Searle, semantic properties, notably intentionality, can be possessed both by mental states such as particular judgements and by linguistic tokens such as inscriptions and illocutionary acts. (Searle 1983, p. 27) An illocutionary act shares some of the semantic properties of the mental state it expresses precisely because it is an expression of that mental state. (Searle 1983, p. 9) This expression takes place when the speaker performs a meaning-conferring act that embues the spoken sounds (or written shapes) with the content of an intentional state that those sounds are intended to express. This is an instance of what Searle calls ‘Intentional causation’ (1983, pp. 122-123): the speaker’s ‘meaning intention’ that the sounds express an intentional state is a *cause* of the fact that the utterance comes to have intentionality. Indeed, in providing a causal explanation of the intentionality of linguistic tokens such as utterances and inscriptions, it seems necessary to invoke two meaning-laden mental states of the speaker (1) the mental state expressed by the speech act, and (2) the intentional (i.e., purposeful) act by which the content of this mental state is conferred upon the sounds uttered.

If the expression ‘derived intentionality’—or more generally, ‘derived *semantic properties*’—is meant to signify this sort of causal dependence, we may clarify the usage of the expression in the following way:

Causally Derived Semantics:

X has semantic property *P* derivatively *iff.*

(1) X has semantic property P

(2) There is some Y such that:

a) $Y \neq X$, and

b) Y has semantic property P

c) Y’s having P is a (perhaps partial) cause of X’s having P.

This notion is applied against CTM in the following way: the semantic properties of symbols are causally derived from those of mental states, whereas the semantic properties of mental states are not derived at all. Therefore it is quite wrong-headed to explain the semantics of mental states by appealing to the semantics of symbols, since (i) the semantics of mental states are not derived at all, but intrinsic, and (ii) any explanation in terms of meaningful symbols would require a further inquiry into the source of the semantic properties of the symbols, which would lead back to the semantic properties of mental states in any case, thus involving the explanation in circularity and regress.

The preceding argument initially seems very persuasive. The argument, however, depends upon the following two claims:

A: Necessarily, all symbols with semantic properties have those properties derivatively.

B: All semantic properties of mental states are underived.

Both of these claims are open to serious question. Undefended, (B) simply begs with question against CTM, since it is one of the basic claims of the theory that the semantic properties of mental states are derived from those of mental representations. Daniel Dennett (1987), moreover, has suggested that the semantic properties of high-level cognitive states such as beliefs and desires of human beings might be derived from lower-level cognitive states, and ultimately from the ‘intentions’ of our genes. While Dennett’s suggestion that subcognitive processes (much less genes) can be bearers of intentional states may seem dubious to many, writers on this subject do seem to experience a very strong and basic clash of intuitions on the issue of whether the semantic properties of mental states could be anything other than intrinsic. As in all cases where the issue seems to rest with clashing intuitions, it is difficult to see a way beyond this impasse.

The objections that have been raised against (A), by contrast, seem easier to assess. For what writers like Searle and Sayre have really argued is something weaker than (A): namely, that certain familiar kinds of symbols—perhaps *all* of the familiar kinds, but notably speech acts, inscriptions, and symbols in computers—have their semantic properties in a fashion that is causally derived. To this point, many advocates of CTM, Fodor included, cheerfully agree. Indeed, Fodor's account of the semantic properties of symbols in overt speech is strikingly similar to Searle's. (Cf. Fodor 1981, p. 31) Where Fodor parts ways with Searle is over the further inference that the *only way any* symbols *could* come by semantic properties is in a fashion dependent upon other meaning-bearing entities or meaning-bestowing acts. So while Searle is right in saying that the semantic properties of speech acts are causally traceable to the intentional states of speakers (and perhaps even that the semantic properties of symbols in production-model computers are similarly traceable to meaning-bestowing acts of programmers and users), Fodor would say that *the symbols of Mentalese differ precisely in this regard*: their semantic properties are not derived, but intrinsic, and Searle has wrongly generalized from properties of some classes of symbols to conclusions about all possible symbols.

This line of response to the causal dependence objection seems correct, so far as it goes. The case that has been made for (A) depends upon the examples of speech acts, inscriptions and symbols in production-model computers. These examples provide at best inductive evidence for (A), but no deductive proof. For the causally derived intentionality objection to succeed, it would be necessary to show that no symbols could have semantic properties intrinsically. And the only way to do this would be to investigate the nature of symbols and symbolic meaning.

4. Conceptually Dependent Intentionality

The Causal Derivation Objection shares with CTM the assumption that there are certain properties called 'intentionality,' and 'semantic properties' that are possessed *both* by mental states *and* by symbols such as utterances and inscriptions. There is, however, another way of viewing the situation: It is undoubtedly the case that the *words* 'intentionality', 'semantics', 'meaning', 'reference', and the rest of the semantic vocabulary are used in connection with mental states and symbols alike. But it does not follow from this that these words comprising the semantic vocabulary *express* the same *properties* in both contexts: that is, there may be a systemmatic ambiguity in the semantic vocabulary so that locutional schemas such as "*A* means...." and "*A* is about...." function in one way when a mental-state-term is substituted for *A*, and function in a different way when a symbol-term is substituted. That is, we *say* both:

(1) Many of John’s thoughts have been about Mary of late.

and

(2) The inscriptions of the name ‘Mary’ in John’s diary are about Mary.

But the expression “___ is/are about Mary” may mean something different (i.e., may express a different property) in (1) from what it means in (2).

The suggestion I wish to develop is that the semantic vocabulary is indeed systemmatically ambiguous. More precisely, semantic terms are what Aristotle called *paronyms* or homonyms with a focal meaning. Aristotle’s classic example of paronymy is in the different yet related uses of the word ‘healthy’. Aristotle points out that many things are *called* “healthy,” but we are making different claims about different things when we call them “healthy.” When we call a *person* “healthy,” for example, what we mean is that she is in *good health*, whereas when we call a *food* “healthy” we mean that it is *conducive to health*, and when we call someone’s *appearance* “healthy” we mean that it is *indicative of health*. This is a special kind of homonymy called *paronymy*, in which one of the meanings (in this case, the one that denotes bodily health) is *focal*, in that the others all refer back to it: healthy food is food that makes for healthy bodies.

The expression ‘derived intentionality’ seems suggestive of the possibility of an idea that might be more perspicuously represented as “derived ‘intentionality’ ”—that is, the view that ‘intentionality’ is a paronymous term, for which which the sense of the word that is applied to symbols is dependent upon (“derivative from”) the sense that is applied to the mind in much the way that the sense of ‘healthy’ that applies to food is dependent upon the sense that applies to living organisms. And as with ‘intentionality’, likewise with the rest of the semantic vocabulary. This may also have been the point behind Sayre’s (1986) claim that:

> Inasmuch as the English word ‘cat’ refers to cats, the word consists of more than can be uttered or written on paper. It consists of the symbolic form CAT (which can be instantiated in many ways in speech and writing) *plus interpretive conventions by which instances of that form are to be taken as referring to cats*. Similarly, the symbolic form GO means the opposite of STOP (or COME, etc) by appropriate interpretive conventions of English, while by those of Japanese it means a board game played with black and white stones. But *without interpretive conventions it means nothing at all*. (1986, p. 123, emphasis added)

If this is the right analysis of symbolic meaning, then the problem is not that we have a notion of meaning and only know how to hook it up with symbols through conventions and intentions. The conventionality comes in, not in the *transmission* or *acquisition* of symbolic meaning, but in the very *property* of symbolic meaning itself. We have several notions of ‘meaning,’ one of which is applicable to mental states, and others to symbols. We may mark this distinction by saying that the *semantic vocabulary can be used to express at least two different properties*: the ‘mental-semantic’ properties of mental states, and the ‘semiotic-semantic’ properties of symbols. The notion of ‘meaning’ that applies to symbols is convention- and intention-dependent to the core, and anyone who says he has a non-conventional notion of 'meaning' for symbols is (whether he knows it or not) no longer using the word 'meaning' in the familiar way, but in some fashion discontinuous with familiar usage.[2] The task for the advocate of a kind of non-conventional, non-intentional 'meaning' for symbols is thus not to show how to hook up the familiar notion of meaning with symbols in a new way; rather, it is first and foremost *to say how he is using the word 'meaning' when he applies it to symbols*, and how that relates to the kind of meaningfulness that is attributed to mental states.

This notion of conceptual dependence may be developed in the following way: *A concept X is conceptually dependent upon a concept Y just in case any adequate analysis of X will include mention of Y*[3]. This may be applied to the semantic vocabulary in the following manner: Words in the semantic vocabulary, such as 'means', express different properties when applied (i) to symbols and (ii) to mental states; and an analysis of the properties expressed by the uses of semantic terms applied to symbols will necessarily refer to the kinds of properties expressed by the uses of those terms as applied to mental states. (E.g., an analysis of symbolic meaning will involve references to 'mental meaning.')

In the next sections of this paper, I shall examine these claims and make a case to the effect that they provide the basis for a strong argument against CTM's account of intentionality. This examination takes us into an brief discussion of semiotics, and it may be easy to lose the main thread of argument, so I shall begin by summarizing the argument which I shall develop in the following sections.

1. Words in the semantic vocabulary (such as 'meaning', 'intentionality' and 'reference') are systemmatically homonymous, having separate usages that apply
   (a) to mental states and
   (b) to symbols.
2. We may refer to the properties picked out by these usages as *mental-semantic* properties and *semiotic-semantic* properties, respectively.
3. The relationship between mental-semantics and semiotic-semantics is not plain homonymy, but *paronymy*: expressions used to attribute semiotic-semantic properties are conceptually dependent upon expressions used to attribute mental-semantic properties.
4. The analysis of attributions of semiotic-semantic properties reveals this dependence because the analysis of attributions of semiotic-semantic properties refers back to the mental-semantic properties of mental states.
5. Any attempt to account for the mental-semantic properties of mental states in terms of the semiotic-semantic properties of symbols (be they 'Mentalese' or otherwise) would be circular and regressive.
6. When CTM speaks of "mental representations having syntactic and semantic properties, the 'semantic properties' it speaks of are either (A) mental-semantic properties" (B) semiotic-semantic properties, or else (C) CTM's advocates are using the expression 'semantic properties' in some third and undisclosed way.

7. It makes no sense to adopt interpretation (A).
8. If we adopt interpretation (B), CTM's account of intentionality is involved in circularity and regress.
9. If we adopt interpretation (C), it is quite unclear what CTM is claiming, as the usage of the key expression 'semantic properties' diverges from familiar pre-existing usage; such a theory could not be assessed until its advocates provided a rule for this new use of the word 'semantic'.
10. On none of these interpretations has CTM provided a viable account of the intentionality of mental states. Indeed, the initial plausibility of the account trades in large measure upon a blindness to the ambiguity of the semantic vocabulary.

5. Analyzing the Semantic Vocabulary

The first claim made by the Conceptual Dependence Objection is that the semantic vocabulary is systemmatically paronymous. This claim is best established by an examination of the semantic vocabulary as it is applied to symbols. I have developed a substantially longer analysis of the notion of symbolic meaning elsewhere, (Horst, 1996) the salient results of which will be presented here. As a preliminary to examining the notion of symbolic meaning, however, it is helpful to clarify the usage of the word 'symbol' in its own right, for its usage is already fraught with dangerous ambiguities. Sometimes the word 'symbol' is used specifically to denote things such as words insofar as they are semantically typed—that is, things that symbolize something. According to this usage, words such as 'dog' are symbols, but graphemes (e.g., <P>—I shall observe the covention of using angle brackets to set off mention of graphemes throughout this paper) and *phonemes* (e.g., /f/) in their own right are not. But the word 'symbol' also has a distinct and broader usage, that applies to tokens of (e.g.) graphemic and phonemic types as well. It is not at all odd, for example, to speak of the letters on an eyechart as 'symbols' even though they have no semantic properties. Both usages appear in the literature on symbols and the mind (e.g., Fodor (1981), pp. 21, 22, 20; Haugeland (1981), pp. 21–22), and so it is useful to sidestep any possible misunderstandings by replacing the word 'symbol' with some technical locutions that do not share this ambiguity. I shall follow the practice of Horst (1996) in using the word 'signifier' for the usage involving semantic typing, and the word 'marker' for the broader usage encompassing graphemic and phonemic typing.[4] Thus there is a marker-type <P> in the Roman alphabet, but that marker type is not employed in its own right for signification in English. There is a marker-sequence <d-o-g> that is permissible under English spelling conventions, and English semantic conventions associate this marker-sequence with a particular meaning to produce the signifier-type 'dog'.

5.1. MARKERS

It should be noted from the outset that marker-types are *conventional* in nature. By this I mean that part of what is involved in saying, e.g., that a particular squiggle is a letter <P> is to locate it within a particular conventionally-established type, and not merely to identify it by its natural features such as shape. Arguably, for example, the same things that could count as (capital) <P>s could also count as (capital) rhos. But being-a-<P> and being-a-rho are not the same thing at all. Better to say that there are two different conventionally-established types, <P> and rho, employed (at least initially) by different linguistic communities. Each of these types involves criteria for what patterns an inscription must have to be able to count as a token of that type, and the criteria for <P> happen to coincide with the criteria for rho.

It is also important to note a further ambiguity that arises in trying to say what kind of marker a given individual is a token of. Suppose that the local optometrist, Dr. Onassis, inscribes the following:

Figure 1. Symbol on Eyechart

Mr. Smith sees it and identifies it as a <P>, but Dr. Onassis says, ‘No, no, it’s not a <P> but a Rho.’ If Dr. Onassis and Mr. Smith are sufficiently obstreperous, they might even proceed to have a heated argument on the subject. Mr. Smith might indignantly contend that he has known how to spot a <P> since he was four years old, and that this is a <P> if ever he has seen one. To which Dr. Onassis might reply that, since he drew the figure, after all, he is in a priviliged position to say what letter it is, and he in fact drew up this eyechart for his substantial constituency of Greek patients.

What lies at the heart of this somewhat crossed-purposed argument is the fact that there are several different things that can be said about the identity of a marker token, all of which tend to be expressed by way of the locutional schema “___ is a ....’ For we might use this expression to talk about:

i) what marker-type a particular squiggle X is conventionally *interpretable* as being a token of (i.e., what marker-type it satisfies the criteria for)
ii) what marker-type it was *intended* to be a token of
iii) what marker-type it was *interpreted* as being a token of, and
iv) how it might, *in principle,* be interpretable as a token of a marker-type.[5]

The confusion between Dr. Onassis and Mr. Smith arises when they assume that there is just one thing called ‘being a <P>’ (or an Rho), when in fact, there are four senses of ‘being’ a token of a marker-type. To express any of these four senses, one might, in ordinary language, *say* ‘X is an M.’ But what *we mean* by such an utterance depends upon whether we are talking about how it is *interpretable,* how it was *intended,* how it was *interpreted* or how it is, *in principle, interpretable.* And depending on which of these things we mean, our utterance of ‘X is an M’ might have any of four different logical analyses:

- ‘X is *interpretable* (under convention C of linguistic group L) as a marker of type M’
- ‘X was *intended* (by its author S) as a marker of type M’
- ‘X was *interpreted* (by some H) as a marker of type M’, and
- ‘X is *interpretable-in-principle* as a marker.’

I wish to follow Horst (1996) here in holding that what we have here is a true disambiguation of the notion of ‘being an M’ where ‘M’ denotes a marker type. That is, these four technical locutions each represent something that could be meant in saying that something is a marker, and jointly exhaust the ordinary meaning of saying that something is, e.g., a <P>. In other words, (a) there is no *one* question of whether something ‘is’ a <P>, but separate questions about how it is interpretable under particular conventions, how it was intended by its author, how it was interpreted by those who apprehend it, and how it is, in principle, interpretable; and, moreover, (b) once one has answered *these* questions, there is no *further* question to be answered about what type of marker it just plain *is*.

5.2. SIGNIFIERS AND SYMBOLIC MEANING

A precisely analogous set of issues arises with respect to signifiers and symbolic meaning. In order for an entity to be a signifier—i.e., a semantically-typed symbol such as the word ‘dog’—two things must be the case. First, it must be a token of a marker type by having a conventionally-sanctioned graphemic or phonemic pattern, or some other corresponding pattern of markers (e.g., dots and dashes in Morse code). Second, this marker-type must be associated with some sort of semantic value—for example, its sense or its denotation. But even if we set aside all controversy about theories of sense and reference, the issue of the ‘meaning’ of a term is yet complicated by a certain amount of ambiguity. Consider, for example, the following inscription:

Pain

Figure 2. Ambiguous Inscription

What is the meaning of what has been inscribed here? We can again imagine a situation in which one person, familiar with English semantic conventions, says that it is a general term indicating a particular unpleasant sensation (i.e., pain), while another person, familiar with French semantic conventions, says that it is a general term indicating bread. It is, in point of fact, possible to generate heated arguments in a classroom about what this inscription ‘really means’ and how one ought to settle it. The answers people give seem to reflect the following four perspectives: (1) that it means ‘pain’ in English *and* means ‘bread’ in French (i.e., that ‘the meaning’ of a term is decided by its *compatibility* with semantic conventions and is indeed *relative* to those conventions), (2) that ‘what it means’ is determined by the intentions of the inscriber, (3) that it ‘means different things to different people' (i.e., that meaning is a matter of individual acts of interpretation), and (4) that the inscription is compatible with indefinitely many interpretations (this generally from people with a little background in formal semantics or coding theory or some such discipline).

Really, though, debates over what determines the 'real meaning' of the inscription are rather pointless, unless perhaps the point is simply to clarify discrepancies in ordinary usage. What seems more important here is that there are at least *four diferent and interesting things one might say* about the semantic properties of a given token, namely:

i) how it is interpretable under particular semantic conventions
ii) what its author intended it to express
iii) how it was interpreted by some individual who apprehended it, and
iv) how it could, in principle, be interpreted.

The analogy with the four ways of 'being' of a marker type should be immediately obvious. For here again there are four senses of 'being' a token of a signifier type, involving *interpretability under a convention, authoring intention, actual interpretation* and *interpretability-in-principle*. And once again, I wish to suggest, it is both possible and desireable to replace ambiguous expressions such as 'means P' with more perspicuous expressions, namely:

....is interpretable (under convention C) of linguistic group L as meaning P
....was intended (by its author A) as meaning P
....was interpreted (by some H who apprehended it) as meaning P
and ....is interpretable-in-principle as meaning P.

Elsewhere, (Horst, 1996) I have suggested the following analyses for these expressions.

Interpretability under a Convention: *An object x may be said to be interpretable as signifying (meaning, referring to) Y iff:*

(1) x is interpretable as a marker of some type T employed by linguistic group L
(2) There is a convention among members of L that markers of type T may be used to signify (mean, refer to) Y.

Authorial Intention: *An object x may be said to be intended (by S) to signify (mean, refer to) Y iff:*

(1) x was produced by some language-user S
(2) S intended x to be a marker of some type T
(4) S believed that there are conventions whereby T-tokens may be used to signify Y
(3) S intended x to signify Y by virtue of being a T-token.

Actual Interpretation: *An objectx may be said to have been interpreted (by H) as significant (meaning, referring to) Y iff:*

(1) Some language user H apprehended Y
(2) H interpreted x as a token of some marker type T
(3) H believed that there to be a linguistic convention C licensing the use of T-tokens to signify Y
(4) H construed x as signifying Y by virtue of being a T-token.

Interpretability in Principle: *An object x may be said to be interpretable-in-principle as signifying Y iff:*

(1) x is interpretable-in-principle as a token of some marker type T
(2) There could be a linguistic community L that employed a linguistic convention C such that T-tokens would be interpretable as signifying Y under convention C.

Or, equivalently,

An object x may be said to be interpretable-in-principle as signifying Y iff:

(1) x is interpretable-in-principle as a token of some marker type T
(2) There is a mapping M available from a set of marker types including T to a set of interpretations including Y
(3) $M(T) = Y$.

5.3. SYNTAX

Finally, the same issues occur at the syntactic level as well. If we ask about the syntactic properties of a string of markers, e.g.:

Fox

Figure 3. Ambiguous String

it quickly becomes evident that the answer will depend upon what symbol-game is operative. This string could have *no* syntactic properties (e.g., if it is a line of an eyechart); it could be a representation of the English word ‘fox’, in which case it has internal structure (which might be called ‘syntactic’ on the dubious assumption that spelling can be subsumed under ‘syntax’) and outward-looking syntactic structure insofar as it is of a particular grammatical class. But it could be a formula in the predicate calculus consisting of a predicate letter and two arguments (‘o’ and ‘x’), or a predicate letter and one argument (‘ox’). In each case this concatenation of markers has very different syntactic structure. If one presses people to say which syntactic structure it ‘really has’ one gets the same classes of answers: (1) that it counts equally as a token of each of the types whose criteria it satisfies, (2) that the issue turns upon the intentions of the inscriber, (3) that it is different things to different people, and (4) that it could, in principle, have an infinite variety of syntactic construals, as there are infinitely many symbol-games in which it could occur. Again, however, there is no point in arguing which of these ‘really determines’ syntactic form. Better once again simply to distinguish questions about how it is interpretable under syntactic rules of particular language-games (e.g., predicate calculus, written English spelling, etc.), from questions about how it was or might be intended or interpreted. But once we do this, we become aware that syntactic categories, being bound up in the use of language-games employed by human communities, are more than abstract combinatorial properties. Syntax, as opposed to mere concatenation, is conventional in nature in much the fashion that marker-typing and semantic-typing are conventional, in that syntactic types such as ‘conjunction sign,’ and ‘count noun’ are established as part of the set of language-games of a linguistic community, and are not strictly reducible to combinatorial properties of markers employed by those communities.

5.4. APPLYING THE SEMIOTIC ANALYSIS TO COMPUTERS

In light of the centrality of the claim that computers are symbol-manipulators, it is curious that virtually nothing has been written about how computers may be said to store and manipulate symbols. While in principle everything said here can be applied to representations in computers, spelling out the details is not a trivial problem from the standpoint of semiotics. Unlike utterances and inscriptions (and the letters and numerals on the tape of Turing’s computing machine), most symbols employed in real production-model computers are never directly encountered by anyone, and most users and even programmers are blissfully unaware of the conventions that underlie the possibility of representation in computers. Spelling out the whole story in an exact way turns out to be cumbersome, but the basic conceptual resources needed are simply those already familiar from the Semiotic Analysis. I shall give a general sketch of the analysis here, and direct the reader who desires a fuller version to the Appendix of Horst (1996).

The really crucial thing in getting the story right is to make a firm distinction between two questions. The first is a question about semiotics:

> *In virtue of what do things in computers count as markers, signifiers and counters?*

The second is a question about the design of the machine:

> *What is it about computers that allows them to manipulate symbols in ways that ‘respect’ or ‘track’ their syntax and semantics?*

Of course, while one could design computers that operate (as Turing’s fictional device did) upon things that are already symbols by independent conventions (i.e., letters and numerals), most of the ‘symbols’ in production-model computers are not of this type, and so we need to tell a story about how we get from circuit states to markers, signifiers and counters. I shall draw upon two examples here:

Example 1: The Adder Circuit:

In most computers there is a circuit called an *adder*. Its function is to take representations of two addends and produce a representation of their sum. In most computers today, each of these representations is stored in a series of circuits called a *register*. Think of a register as a storage medium for a single representation. The register is made up of a series of 'bistable circuits'—circuits with two stable states, which we may conventionally label '0' and '1', being careful to remember that the numerals are simply being used as the *labels* of states, and are not the states themselves. (Nor do they represent the numbers zero and one.) The states themselves are generally voltage levels across output leads, but any physical implementation that has the same on/off properties would function equivalently. The adder circuit is so designed that the pattern that is formed in the output register is a function of the patterns found in the two input registers. More specifically, the circuit is designed so that, under the right interpretive conventions., the pattern formed in the output register has an interpretation that corresponds to the sum of the numbers you get by interpreting the patterns in the input registers.

Example 2: Text in Computers

Most of us are by now familiar with word processors, and are used to thinking of our articles and other text as being 'in the computer,' whether 'in memory' or 'on the disk.' But of course if you open up the machine you won't see little letters in there. What you will have are large numbers of bistable circuits (in memory) or magnetic flux density patterns (on a disk). But there are conventions for encoding graphemic characters as patterns of activity in circuits or on a disk. The most widely used such convention is the ASCII convention. By way of the ASCII convention, a series of voltage patterns or flux density patterns gets mapped onto a corresponding series of characters. And if that series of characters also happens to count as words and sentences and larger blocks of text in some language, it turns out that that text is 'stored' in an encoded form in the computer.

Now to flesh these stories out, it is necessary to say a little bit about the various levels of analysis we need to employ in looking at the problem of symbols in computers and also say a bit about the connections between levels. At a very basic level, computers can be described in terms of a mixed bag of *physical properties* such as voltage levels at the output leads of particular circuits. Not all of these properties are related to the description of the machine as a computer. For example, bistable circuits are built in such a way that small transient variations in voltage level do not effect performance, as the circuit will gravitate towards one of its stable states very rapidly and its relations to other circuits are not affected by small differences in voltage. So we can idealize away from the properties that don't matter for the behavior of the machine, and treat its components as *digital*—i.e., as having an integral and finite number of possible states.[6] It so happens that most production-model computers have many components that are *binary*—they have *two* possible states—but digital circuits can, in principle, have any (finite, integral) number of possible states. Treating a machine that is in fact capable of some continuous variations as a digital machine involves some idealization, but then so do most descriptions relevant for science. The digital description of the machine picks out properties that are *real* (albeit idealized), *physical* (in the strong sense of being properties of the sort studied in physics, like charge and flux density) and non-conventional.

Next, we may note that a *series* of digital circuits will display some *pattern* of digital states. For example, if we take a binary circuit for simplicity and call its states ‘0’ and ‘1’, a series of such circuits will display some pattern of 0-states and 1-states. Call this a **digital pattern**. The important thing about a digital pattern is that it occupies a level of abstaction sufficiently removed from purely physical properties that the same digital pattern can be present in *any* suitable series of digital circuits independent of their physical nature. (Here ‘suitable series’ means any series that has the right length and members that have the right number of possible states.) For example, the same binary pattern (i.e., digital pattern with two possible values at each place) is present in each of the following sequences:

a a b b

0 0 1 1

• • | |

Figure 4. Digital patterns of graphemes.

It is also present in the music produced, by playing either of the following:

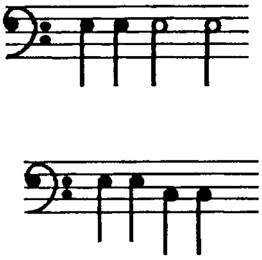

Figure 5. Digital patterns of notes.

And it is present in the series of movements produced by following these instructions:

(1) Jump to the left, *then*
(2) jump to the left again, *then*
(3) pat your head, *then*
(4) pat your head again.

Or, in the case of storage media in computers, the same pattern can be present in any series of binary devices if the first two are in whatever counts as their 0-state and the second two are in whatever counts as their 1-state. (Indeed, there is no reason that the system instantiating a binary pattern need be physical in nature at all.)

Digital patterns are *real*. They are *abstract* as opposed to *physical* in character, although they are literally present in physical objects. And, more importantly, they are *non-conventional*. It is, to some extent, our conventions that will determine which abstract patterns are important for our purposes of description; but the abstract patterns themselves are all really there independent of the existence of any convention and independently of whether anyone notices them.

It is digital patterns that form the (real, non-conventional) basis for the tokening of symbols in computers. Since individual binary circuits have too few possible states to encode many interesting things such as characters and numbers, it is *series* of such ciruits that are generally employed as units (sometimes called 'bytes') and used as symbols and representations. The ASCII convention, for example, maps a set of graphemic characters to the set of 7-digit binary patterns. Integer conventions map binary patterns onto a subset of the integers, usually in a fashion closely related to the representation of those integers in base-2 notation.

Here we clearly have conventions for both markers and signifiers. The marker conventions establish kinds whose physical criterion is a binary pattern. The signifier conventions are of two types. In cases like that of integer representation, we find what I shall call a *representation scheme*, which directly associates the marker type (typified by its binary pattern) with an interpretation (say, a number or a boolean value). In the case of ASCII characters, however, marker types typified by binary patterns are not given semantic interpretations. Rather, they *encode* graphemic characters that are employed in a pre-existing language-game that has conventions for signification; they no more have meanings individually than do the graphemes they encode. A string of binary digits in a computer is said to 'store a sentence' because (a) it *encodes* a string of characters (by way of the ASCII convention), and (b) that string of characters is used in a natural language to express or represent a sentence. I call this kind of convention a *coding scheme*. Because binary strings in the computer encode characters and characters are used in text, the representations in the computer inherit the (natural-language) semantic and syntactic properties of the text they encode.

It is thus clear that computers can and do store things that are intepretable as markers, signifiers and counters. On at least some occasions, things in computers are intended and interpreted to be of such types, though this is more likely to happen on the engineer's bench than on the end-user's desktop. It is worth noting, however, that in none of this does the computer's nature *as a computer* play any role in the story. The architecture of the computer plays a role, of course, in determining what kinds of resources are available as storage locations (bistable circuits, disk locations, magnetic cores, etc.). But what makes something in a computer a symbol

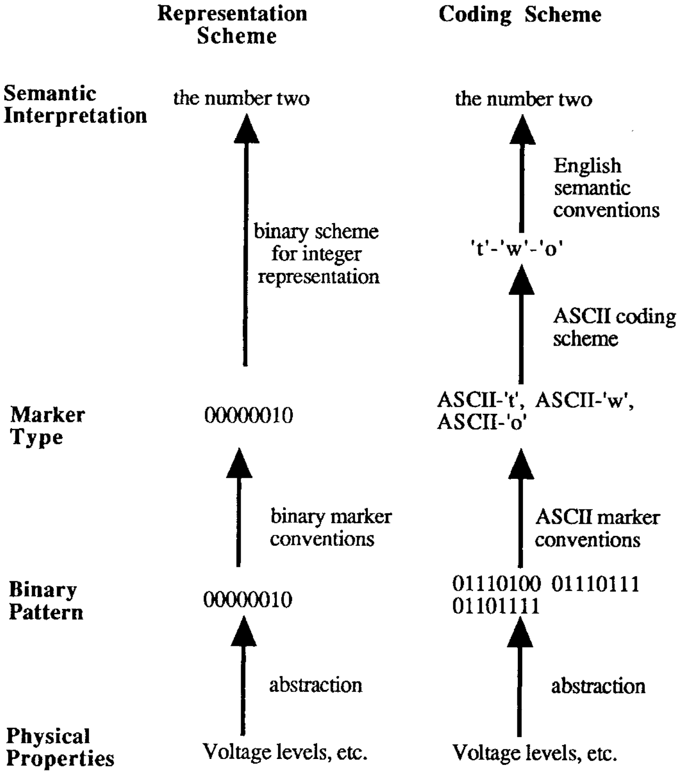

Figure 6. Representation and Coding Schemes

(i.e., a marker) and what makes it meaningful are precisely the same for symbols computers as for symbols on paper: namely, the conventions and intentions of symbol-users.

Now of course the difference between computers and paper is that computers can do things with the symbols they store and paper cannot. More precisely, computers can produce new symbol-strings on the basis of existing ones, and can do so in ways that are useful for enterprises like reasoning and mathematical calculation. The common story about this is that computers do so by being sensitive to the syntactic properties of the symbols. But strictly speaking this is false. Syntax, as we have seen, involves more than functional description. It involves convention as well. And computers are no more privy to syntactic conventions than to semantic ones. For that matter, computers are not even sensitive to *marker* conventions. That is, while computers operate upon entities that happen to *be* symbols, the computer does not relate to them *as* symbols (i.e., as markers, signifiers and counters). To do so, it would need to be privy to conventions.

There are really two quite separate descriptions of the computer. On the one hand, there is a functional/causal story; on the other a semiotic story. The art of the programmer is to find a way to make the functional/causal properties do what you want in transforming the symbols. The more interesting symbolic transformations you can get the functional properties of the computer to do for you, the more money you can make as a computer programmer. So for a computer to be *useful*, the symbolic features need to line up with the functional/causal properties. But they *need not in fact* line up, and when they do it is due to an excellence in design and not to any *a priori* relationship between functional description and semiotics.

6. The Paronymy of the Semantic Vocabulary

We have now established that a certain amount of ambiguity exists in ordinary attributions of semantic properties to symbols. As of yet, however, we have said nothing about the relationship between attributions of semantic properties to symbols and attributions of semantic properties to mental states. Nor have we said anything to establish the existence of *paronymy* as opposed to simple homonymy or ambiguity, as indeed there arguably is no paronymy among the different senses of 'being' of a marker-type or a signifier-type. But if one accepts the preceding analysis of attributions of semantic properties to symbols, it quickly becomes apparant (1) that none of the uses of the semantic vocabulary that apply to symbols can be the same as the usage that applies to mental states, and (2) that the former are conceptually dependent upon the latter.

Consider the analyses presented for what might be meant in saying that a particular inscription 'means pain.' This might mean any of four things: (i) that it is interpretable, under particular conventions of a particular linguistic group, as meaning pain, (ii) that it was intended by its inscriber to mean pain, (iii) that it was interpreted by someone as meaning pain (and thus 'meant pain to her') or (iv) that it is, in principle, possible to assign pain to it as an interpretation. A more detailed analysis of each of these four attributions of meaning has revealed that each of them involves reference either to conventions or to meaningful mental states. To say that an inscription was *intended* as meaning pain or *interpreted* as meaning pain is to refer to the intentional states of the author or the interpreter, respectively, and is thus to allude to meaningful mental states. To say something is *interpretable* as meaning pain is at least covertly to allude to the semantic conventions of some linguistic group, and to say that something is *interpretable-in-principle* as meaning pain is to refer to conventions under the modality of possibility. And since conventions involve, at very least, shared beliefs and practices, invoking interpretability indirectly might be analyzed in terms of shared beliefs.

Two observations seem to follow fairly straightforwardly. First, it seems immediately evident that this hidden complexity of attributions of semantic properties to symbols is *not* shared by attributions of semantic properties to mental states. To say that someone's *thoughts* are 'about pain' is *not* to say anything about how those thoughts are conventionally interpretable, or about any authoring intentions or external interpretations, or about the mere availability of an interpretation scheme (i.e., a mapping onto interpretations). The logical form of attributions of semantic properties to mental states does not have the kind of hidden complexity we found in the semiotic cases, but rather seems to stay close to what is suggested by the surface structure. Indeed, the predicates expressed by the semantic vocabulary when it is applied to symbols have *different numbers of arguments* from the predicate expressed when it is applied to mental states. But if this is the case, the semantic verbs express different properties when their grammatical subjects are nouns denoting mental states from those they express when their grammatical subjects are nouns denoting symbols. In short, the semantic vocabulary is homonymous. And we may mark this homonymy by distinguishing the *mental-semantic* properties attributed to mental states from the *semiotic-semantic* properties attributed to symbols.

A second observation is also forthcoming. We have already noted that the analysis of semiotic-semantic properties involved mention of meaningful mental states —namely, those involved in the the shared beliefs constitutive of conventions, those involved in authoring intentions and those involved in acts of interpretation. It thus follows that semiotic-semantic properties are conceptually dependent upon mental-semantic properties in the fashion required for paronymy, since their analysis refers back to states with mental-semantic properties, much as the analysis of ‘healthy’ as applied to foods refers back to the notion of bodily health. We have thus established the first four theses of the argument outlined above in Section 4. What remains is to apply this towards a critique of CTM’s account of the intentionality of mental states.

7. Re-Interpreting CTM

Upon returning to CTM, the first problem that confronts us is one of finding an appropriate way to re-interpret the claims made by the theory, since both the articulation of the theory and much of its intuitive appeal seem to depend upon the assumption that expressions such as ‘meaningful’, ‘intentionality’, ‘representation’, ‘content’, and ‘semantic properties’ may be applied univocally (a) to discursive symbols such as inscriptions and utterances, (b) to mental states such as beliefs and desires, and (c) to the hypothesized ‘symbols’ of Mentalese. And this assumption would seem to ground the intuitive appeal of CTM: if we assume that there is just one class of things called ‘semantic properties,’ which can be possessed both by symbols and by mental states, it is at least initially plausible to suppose that the ultimate locus of these properties might be symbols in a language of thought, and that the semantic properties of mental states are ‘inherited’ from those of Mentalese symbols. Semantic properties, at any rate, seem like the right sorts of things to account for other semantic properties!

Once having seen the paronymy of the semantic vocabulary, however, we are forced not only to re-assess the plausibility of CTM’s account of intentionality, but to *re-examine just what it might be claiming*. For it will no more do to say that symbols and thoughts ‘both have semantic properties such as meaning’ that it will do to say that living bodies, foods and complexions ‘can all be healthy.’

And since the semantic vocabulary denotes different properties when applied to symbols and to mental states, it is no longer clear that the 'semantic' properties of the hypothesized 'symbols' of Mentalese are even the right sorts of things to be potential explainers of the mental-semantic properties of mental states. On the assumption that the vocabulary was univocal, meaningful symbols seemed like potential explainers, because the kind of 'meaning' needed for the mental states was already in the picture. Now it is necessary to explain how mental-semantics enters the picture at all. It is therefore essential to examine two questions:

1) What kinds of properties are the 'semantic properties' attributed to mental representations by CTM?
2) Are these properties of such a sort to be potential explainers of the mental-semantic properties of mental states?

To the first question there are three possible answers, given the foregoing analysis. Either (a) the 'semantic properties' of mental representations are mental-semantic properties, or (b) they are semiotic-semantic properties, or (c) they are some third and distinct kind of properties. Given the stress placed upon the use of the words 'symbol' and 'representation' by CTM's advocates, it might seem obvious and indeed almost obligatory to interpret these writers as attributing to mental representations the kinds of 'semantic properties' normally attributed to symbols—that is, semiotic-semantic properties. If not, it is hard to see why the fuss that is made over these notions. Moreover, if they mean to use the semantic vocabulary in *some other way*, it seems a peculiar and unfortunate oversight on their part not to have warned us, as we are all likely to read them as meaning the same thing by 'symbolic representation' that the rest of us mean. However, it also seems clear that CTM's advocates did not take seriously the possibility that the semantic vocabulary might be homonymous, and hence it is worthwhile to explore the various avenues they might pursue in dealing with the problems presented at this juncture.[7]

The first possibility, that the 'semantic properties' of mental representations are mental-semantic properties, is quickly dismissed. Once we have made the distinction between mental- and semiotic-semantic properties, it is simply very difficult to see what sense could be made of saying that mental representations may be said to 'have semantic properties,' not in the sense that symbols are said to have them, but in the sense that mental states are said to have them. It remains to examine the other two possibilities.

8. Semiotic-semantics, Circularity and Regress

The most natural way of reading CTM's account of the intentionality and semantics of mental states is to read it as the claim that the mental-semantic properties of mental states are to be explained in terms of the semiotic-semantic properties of mental representations. In order to proceed here, it might be helpful to adapt Fodor's own characterization of cognitive states in *Psychosemantics*:

> *The Nature of Propositional Attitudes*
>
> For any organism *O*, and any attitude *A* toward the proposition *P*, there is a ('computational'/'functional') relation *R* and a mental representation *MP* such that
>
> *MP* semiotically-means that *P*, and
>
> *O* has *A* iff *O* bears *R* to *MP*. (Fodor (1987), p. 17, underscoring marks my emendation)

However, this formulation is not yet adequately perspicuous. For there are several distinct senses in which a symbols may be said to '(semiotically-)mean that *P*'—senses corresponding to interpretability under a convention, authorial intention, actual interpretation, and interpretability-in-principle. So there are at least four ways of reading the above claim:

> *Authoring Intention Version:*
>
> For any organism *O* and any cognitive attitude *A* towards a proposition *P*, there is a relation *R* and a mental marker *MP* such that
>
> *MP* was intended as signifying (that) *P*, and
>
> *O* has *A* iff *O* bears *R* to *MP*.
>
> *Actual Interpretation Version:*
>
> For any organism *O* and any cognitive attitude *A* towards a proposition *P*, there is a relation *R* and a mental marker *MP* such that
>
> *MP* was actually interpreted as signifying (that) *P*, and
>
> *O* has *A* iff *O* bears *R* to *MP*.
>
> *Interpretability Version:*
>
> For any organism *O* and any cognitive attitude *A* towards a proposition *P*, there is a relation *R* and a mental marker *MP* such that
>
> *MP* is interpretable under convention *C* as signifying (that) *P*, and
>
> *O* has *A* if *O* bears *R* to *MP*.
>
> *Interpretability-in-Principle Version:*
>
> For any organism *O* and any cognitive attitude *A* towards a proposition *P*, there is a relation *R* and a mental marker *MP* such that
>
> *MP* is interpretable-in-principle as signifying (that) *P*, and
>
> *O* has *A* iff *O* bears *R* to *MP*.

Despite the range of possible interpretations for CTM here, none if them is viable as an account of the semantics of mental states. Perhaps many readers will find this conclusion too obvious to require discussion, but given the problems with previous attempts to disqualify CTM on related grounds, it seems worthwhile to develop the argument. To do so, I shall deal with the first three versions together, then deal with the interpretability-in-principle version separately. These critiques are developed more thoroughly in Horst (1996).

8.1. CONVENTIONS, INTENTIONS AND CTM

There are several reasons why semiotic-semantic properties of mental representations involving conventions or intentions (i.e., the Interpretability, Authoring Intention and Actual Interpretation Versions) could not provide an explanation of the mental-semantic properties of mental states. First, it is highly doubtful that there are such conventions, authorial intentions or actual acts of interpretation underlying mental states. An account that relied on the existence of such conventions and intentions would thus be highly doubful on empirical grounds. Second, the question of whether there are such conventions and intentions (say, on the part of sophisticated Martian psychologists with brain probes) is wholly irrelevant to the question of what those mental states mentally-mean: what our mental states are about is not affected in the least by the presence or absence of Martian interpreters. The issue of conventional or actual interpretations for brain states is of no importance in deciding the question of what my thoughts are about.

Moreover, explaining mental-semantic properties of mental states in a fashion that depends upon the conventional, intended or interpreted senses of symbols results in explanations that are circular and regressive. They are circular because the mental-semantic properties of mental states would be explained in terms of the semiotic-semantic properties of mental representations, while the semiotic-semantic properties of any symbol must, by definition, be explained in terms of the mental-semantic properties of mental states. They are regressive because the explanation of the mental-meaning of a mental state M is given in terms of the semiotic-meaning of a representation R, which must in turn be explained in terms of the mental-meaning of some further mental state M^*, and so on ad infinitum. In short, one cannot build a viable theory of the mental-semantics of mental states in terms of semiotic-semantic interpretability, intention or actual interpretation.

8.2. SEMIOTIC-SEMANTIC INTERPRETABILITY-IN-PRINCIPLE

There are likewise several problems with the version of CTM based upon semiotic-semantic interpretability-in-principle. I shall discuss two of these, one familiar and one not.[8] First, it is by now notorious (a) that *individual* symbols may be given any interpretation whatsoever, and (b) that even maximally large consistent systems of symbols are susceptible to multiple interpretations, including number-theoretic interpretations. It follows from this that semantic interpretability-in-principle is not a strong enough notion to yield the kind of detemilnacy we find in mental-meaning of mental states. If thought involves representations in a language of thought, the representation formed when I think about pain is interpretable-in-principle as being about anything whatsoever. And even if we restrict interpretability-in-principle to an entire language of Mentalese, the representation formed has a number-theoretic interpretation. Yet my thought mentally-means pain and does not mentally-mean some number-theoretic entity or proposition. Thus semiotic-semantic interpretability-in-principle for representations cannot be a sufficient condition for determining the mental-semantic properties of a mental state.

There is also a second sort of reason why this strategy will not work. Semiotic-semantic properties are only defined over *markers*. And markers *are conventional in nature*. That is, it is not only the *semantic* properties of symbols that are conventional in nature, but also the very fact of *being* a symbol in the bare sense of being a marker. Recall that this is why being-a-<P> and being-a-Rho are two different things, even though the two marker-types share the same spatial criteria. Thus, explaining mental-semantics in terms of the properties of mental symbols involves problems about conventions at the marker level even if we avoid these problems at the *semantic* level by talking about interpretability-in-principle rather than conventions or intentions of speakers and hearers. And here the same problems of circularity and regress will re-emerge.

It thus seems that the kinds of ‘semantic properties’ normally attributed to symbols—i.e., semiotic-semantic properties—cannot plausibly be applied to mental representations, and even if they could, this would not provide the basis for a viable theory of the mental-semantics of mental states. If we are to find an acceptable construal of CTM, we must look elsewhere.

9. An Independent Semantics for Representations

The argument thus far has been that, if the words ‘symbol’, ‘syntax’ and ‘semantics’ express the same properties when applied to mental representations that they express when applied to garden-variety symbols, then CTM’s account of intentionality fails, and with it the attempt to vindicate intentional psychology. And indeed, if the semiotic vocabulary is interpreted in this fashion, it is hard to see the whole theory as anything but deeply confused. This is, I believe, the very result foreseen by Searle and Sayre. From this, however, they seem also to draw the stronger conclusions (1) that there is no way of interpreting the semiotic vocabulary that avoids this kind of confusion, and (2) that as a result, the computational paradigm is bankrupt as a research strategy for a *psychology* of cognition. I wish to resist these further conclusions. In the remainder of this paper, I shall suggest that there is another basic way of interpreting the language used in CTM that avoids the incoherence that results from ascribing semiotic-semantic properties to representations, but only at the significant cost of failing to provide an account of intentionality or a vindication of intentional psychology. Nonetheless, such a version of CTM may very well provide a useful research framework for empirical psychology even if it fails to produce these two distinctively *philosophical* results.

9.1. SYMBOLS, SYNTAX AND MACHINE-COUNTERS

Before addressing the question of *semantics*, however, it is necessary to address the topics of *syntax* and *symbolhood* first. If the ordinary uses to these terms express convention-dependent properties, it nonetheless seems that discussions of computers at least sometimes are really after something else: properties that are both formal and non-conventional. If we can bring such a notion to adequate clarity, it will not really matter that it does not share all of the features of symbolhood (i.e., markerhood) and syntax. I believe that we may do this by beginning with the notion of a functionally-describable device, and then identifying the relevant notions in terms of states of that device. Intuitively, we may define a *machine-counter* as anything that plays the kind of functional role that a symbol was supposed to play in a Turing machine or other functionally-defined symbol-manipulator. More formally,

A tokening of a machine-counter of type T may be said to exist in C at time t iff:

(1) C is a digital component of a functionally-describable system F.

(2) C has a finite number of determinable states $S : \{s_1, \ldots s_n\}$ such that C's causal contribution to the functioning of F is determined by which member of S C is in.

(3) Machine-counter type T is constituted by C's being in state s_i, where $s_i \in S$.

(4) C is in state s_i at t.

The notion of a machine-counter seems suitable for expressing all of the formal and functional properties relevant to 'symbol-processing' in a fashion that avoids references to conventions or intentions. Functions over machine-counters are symbol-manipulations minus the symbols. Symbol-manipulations are functions over machine-counters where the machine-counters have semiotic interpretations.

9.2. SEMANTICS FOR MENTAL REPRESENTATIONS

It remains to consider semantics. What non-conventional properties might one mean to express in saying that mental representations have 'meanings' or 'intentionality'? How might one provide a rule for the use of words like 'meaning' as applied to mental representations? There are, I think, two basic ways of supplying a rule for the use of these words: (1) stipulative definition and (2) theoretical definition. Here I shall confine myself principally to a consideration of the latter. For while there are many possible stipulative definitions of the semiotic vocabulary as applied to representations, stipulation does not seem to be in keeping with the character of discussions of meaning in cognitive science. (Though perhaps some of the characterizations in Newell and Simon (1977) may be an exception.) One could, for example, take some theory of content for representations and use it as a stipulative definition of what 'means' means when applied to representations, but this seems contrary to the widespread assumption that the work done by a theory of representations is independent of the theory of content one gives for those representations. Moreover, it will turn out that the merits of a theory of content will be relevant in just the same ways regardless of whether we use that theory to *define* ‘meaning’ for representations or merely to *explain* it.

The alternative is to construe the semantic vocabulary as applied to representations as a truly theoretical vocabulary: the MR-semantic properties of a representation R that is involved in an intentional state I are those properties, whatever they turn out to be, that account for the mental-semantic properties of I. So if I is a perceptual gestalt of a cat, and it involves a machine-counter (a representation) R, then the MR-meaning of R consists in *those properties of* R, *whatever they are, that are responsible for I’s mentally-meaning ‘cat’*. This formulation seems to meet two important desiderata: it avoids the convention- and intention-dependence of semiotic-semantic properties, and it preserves the independence of a notion of representation and representational ‘meaning’ from specific theories of content.

10. Assessing the MR-Semantic Version of CTM

While this re-interpretation of CTM avoids the problems of conventional regress that afflicted the earlier interpretation, CTM is by no means safely home to port. For a closer inspection reveals that much of the *persuasive force* that was originally enjoyed by CTM was predicated upon the assumptions either (a) that the kind of ‘meaning’ possessed by mental *representations* was precisely the kind possessed by *mental states*, or (b) that the kinds of ‘syntactic’ and ‘semantic’ properties possessed by *representations* was of the kind possessed by symbols. But if MR-semantic properties are theoretically-defined properties distinct from semiotic-semantic and mental-semantic properties, these assumptions are false. With these assumptions undercut by the preceding analysis, it turns out that CTM requires significant justification.

10.1. TWO INTERPRETATIONS OF MR-SEMANTICS

We have characterized MR-semantic properties as ‘those properties of mental representations, whatever they turn out to be, that account for the mental-semantic properties of mental states.’ The hope is that we can use such properties in a larger explanation that would have the following schematic form:

Mental state M has mental–semantic property P because

(i) M involves a relationship to a mental representation MR, and
(ii) MR has MR–semantic property X

Yet there are two ways of substituting our definition of MR-semantics into this schema: a *de dicto* interpretation and a *de re* interpretation. The *de dicto* substitution simply replaces the expression ‘MR-semantic property X’ with its theoretical definition as follows:

De Dicto Interpretation:

Mental state M has mental–semantic property P because

(i) M involves a relationship to a mental representation MR, and
(ii) MR has that property of MR, whatever it is, that accounts for mental-semantic property P.

This, however, yields a pseudo-explanation of a familiar type: on this reading, MR-semantic properties function in this explanation as dormative virtues. The *de re* interpretation, however, is somewhat more promising:

De Re Interpretation

Mental state M has mental–semantic property P because

(i) M involves a relationship to a mental representation MR,
(ii) MR has some property X,
(iii) the fact that MR has X explains the fact that M has P
(iv) X is called an 'MR-semantic property' because
a) it is a property of a mental representation, and
b) it is the property that explains the fact that M has P.

This formulation avoids dormative virtues. It also clarifies what CTM, in and of itself, does and does not provide. First, strictly speaking, CTM does not, in and of itself, provide an explanation of the semantic properties of mental states. It would do so only in conjunction with some other theory that would (1) clarify what these MR-semantic properties are, (2) show how mental representations come to have them, and (3) show how their presence accounts for the presence of mental-semantic properties of mental states. Second, it is important to note how strongly the *persuasive* strength of CTM depends upon the assumption that the 'semantic properties' of representations are the same 'semantic properties' possessed by mental states. If we make this assumption, we may take it for granted that the 'semantic properties' of representations are at least reasonable candidates for explaining the 'semantic properties' of mental states, because there is no special problem of how 'semantic properties' *come onto the scene* when we get to mental states, but only a problem of how they might be *inherited* from things that already have them (i.e., representations). But if we *distinguish* mental-semantic properties from some mysterious property X of mental representations, we are left with the further task of explaining how mental-semantic properties come onto the scene at all. That is, we must show that this X is even a proper *candidate* for explaining mental-semantics and intentionality. If the talk about 'meanings explaining meanings' really means only that there might be some non-intentional properties that could explain meanings, things are not so clear. It is an open question whether any such non-intentional properties are even potential explainers of meaning, and so CTM's ability to say anything about intentionality—even supplemented by an additional semantic theory—turns on a fairly contentious assumption. Perhaps it is an assumption that will be borne out by further investigation—hence the current interest in 'theories of content'—but CTM's ability to contribute to an explanation of intentionality stands or falls with such research.

10.2. THE VINDICATION OF INTENTIONAL PSYCHOLOGY

It is likewise important to see how the foregoing discussion undercuts CTM's claim to vindicate intentional psychology. That vindication turned upon the claim that the computer paradigm had shown us how to coordinate semantic value with causal role by way of syntax. But in light of our distinctions, what we should really say is that the computer paradigm shows that the (convention- and intention-dependent) *semiotic*-semantic properties of symbols in computers can be coordinated with their causal roles. What we need to show to vindicate intentional psychology, however, is something different from this. We must show, first, that the *MR*-semantic properties of representations can be coordinated with causal role, and second, that this would be sufficient for assuring that the mental-semantic properties of mental states could be thus coordinated as well. In brief, the task of coordinating mental-semantic properties with causal role is thus more complicated than that of coordinating semiotic-semantic properties with causal role in a computer:

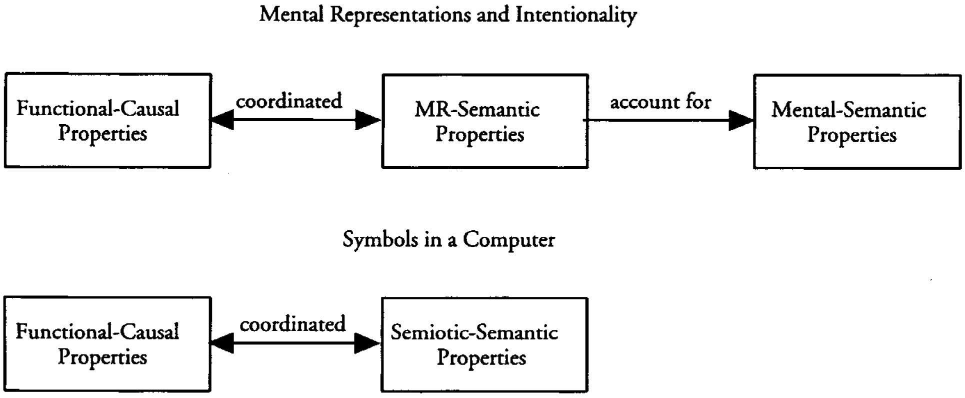
Mental Representations and Intentionality
Functional-Causal Properties
coordinated
MR-Semantic Properties
account for
Mental-Semantic Properties
Symbols in a Computer
Functional-Causal Properties
coordinated
Semiotic-Semantic Properties

Figure 7. Mental Representations and Intentionality

This presents two non-trivial problems. First, we have no proof (from the computer paradigm or anywhere else) that there are *non*-conventional 'semantic' properties mental representations might have that can be properly coordinated with causal role. In particular, conventional symbols avoid the problem of having multiple possible interpretations precisely because meanings are stipulated by convention. It is not clear that one could get the same requisite degree of determinacy without the artifice of stipulation. Second, even if one could view the mind or brain as a computer, there is a considerable problem of then showing that the properties possessed by representations are sufficient to account for the meaningfulness of mental states. The vindication of intentional psychology in this fashion depends upon the ability to find non-intentional properties of representations that explain the mental-semantic properties of mental states. The seriousness of this second problem has, I think, been underestimated in many recent discussions of theories of content.

The issues may be clarified by making use of two distinctions between different kinds of accounts of meaning. First, we may distinguish accounts that explain *what it is to be an X* from those that merely provide a *demarcation* of X's from non-X's. Second, we may also distinguish between the problem of meaning-*assignment* and the problem of meaning*fulness*, and to the corresponding difference between (a) accounts that are concerned with the difference between meaning-X and meaning-Y, and (b) accounts that are concerned with the difference between meaning *something* and meaning *nothing*.

10.3. DEMARCATION AND EXPLANATION

First, it should be a familiar fact that one can provide a materially adequate condition for a predicate without thereby explaining what it is to have that predicate apply to an object. Consider, for example, Aristotle's characterization of humans as 'featherless bipeds.' This definition does not in fact provide even materially necessary or sufficient conditions for being human; but even if it did (due to extinction of apes, bald chickens, amputees, etc.) it would at most provide a criterion for the demarcation of humans from non-humans without an explanation of what it is to be human. Similarly, 'simplest closed two-dimensional polygon' demarcates triangles from non-triangles without telling what it is to be a triangle, whereas 'three-sided polygon' accomplishes the latter as well. I shall mark this distinction by speaking of 'demarcation accounts' and 'conceptually adequate explanations.' (CAEs) In order for X to be a CAE of Y, it must be the case that an ideal understanding of X would itself be enough to derive the conceptual content of Y. Thus, for example, a specification of the laws governing collisions of gas molecules could provide an CAE of the gas laws, since an adequate understanding of large numbers of particle collisions would allow one to derive the gas laws. Reduction is a robust kind of CAE; a weaker sort is the notion of 'instantiation analysis' articulated in Cummins (1983. pp. 17–26).

10.4. MEANING-ASSIGNMENT AND MEANINGFULNESS

Second, we may once again differentiate between two different issues with which a theory of intentionality should be concerned. One question is about what distinguishes things that mean-X from that mean-Y—say, thoughts about horses and thoughts about cows. If one believes that representations are the locus of content, this becomes a question about the difference between representations that 'mean 'horse" and those that 'mean 'cow'.' And there are various approaches to making this distinction, such as Fodor's suggestion that 'horse' tokens are, roughly, those caused by horses. My intention here is not to enter into the already large literature attacking, defending and refining this or other accounts, but merely to make a crucial observation: what this literature addresses is the question of whether particular applications of the notion of causal covariation can produce the right *assignments of meanings* to representations, *given* that they mean *something*. What it tends *not* to address is a second and prior question: namely, What *distinguishes things that mean something from those that mean nothing at all?*

I suggest that we mark this distinction by speaking of the 'problem of meaning-assignments' and the 'problem of meaningfulness.'[9] I take it that CTM originally undertook to provide an account of the meaning*fulness* of mental states—namely, that they inherit their meanings from the representations they contain. Subsequent work on a theory of content for representations seems to be oriented towards assessing its adequacy as a *demarcation* account for meaning-*assignments*. What the distinction between mental-semantic and MR-semantic properties seems to bring forcefully to the fore, however, is the question of whether *we can* have a CAE of mental-semantic properties—that is, a CAE of their meaning*fulness*—in some non-intentional terms. Arguably, for example, notions like causal covariation just do not possess the right conceptual content to provide a CAE of the meaningfulness of mental states, even if they can be made to produce a demarcation criterion for meaning-assignments. It is this issue, I think, that has troubled philosophers who believe that there is an indefeasibly *experiential* or *first-person* or *phenomenological* aspect to mental-intentionality, and it is an issue that has received inadequate attention in the literature.

This, I think, provides a non-trivial challenge for those who wish CTM to provide an account of the intentionality of mental states. First, they must make clear what sorts of properties these MR-semantic properties of representations are supposed to be. Second, they must provide reason to suppose that these properties are adequate to the task of explaining why mental states involving these representations would thereby mentally-mean *something* rather than not mentally-meaning *anything*.

11. Effects of the Conventionality of Syntax

In addition to its proposed explanation of intentionality, CTM is also notable for its alleged ability to account for the productivity of thought by way of syntactic composition of meaningful lexical units. The foregoing considerations have been directed at the viability of CTM's proposals for accounting for the intentionality of the mental insofar as that depends upon the semantic properties of symbols in Mentalese that are lexical primitives. But even if CTM could find an account of MR-semantics for lexical primitives that could avoid the problems already mentioned, my account of semiotics raises additional problems with respect to the semantic properties of *complex* symbols in Mentalese. According to Fodor's account, and indeed any plausible account of a language of thought, there is a finite stock of semantic primitives in Mentalese. It is these to which the ‘causal covariation’ theory of content is supposed to apply. The vast majority of our representations, however, are not primitives, but syntactically-structured complexes whose semantic value is determined by the semantic values of the primitives, in conjunction with syntactically-based compositional rules.

The only notion of ‘syntax’ we have, however, is that which applies to garden-variety symbols employed in a language-game, and that notion is completely convention-dependent. It should be obvious that the ‘syntactic’ rules of the hypothesized language of thought had best not be convention-dependent, or else problems cognate with those developed for semantics will arise. But if some other use of the word ‘syntax’ is at work in CTM, it needs to be developed. Now it might be claimed that the notion of ‘syntax’ at work can be developed completely in combinatorial terms. But this, I think, is unsatisfactory for two reasons. First, what one would then have could not really properly be called ‘syntax,’ because syntax involves more than combinatorial properties. (Combinatorial properties, after all, apply to arrangements of objects generally, and not all arrangements are syntactic arrangements.) Second, what would be lost here in particular is a way of getting a specific semantic value for a complex symbolic structure out of semantic values of its atoms plus combinatorial rules.

The problem might be articulated in the following fashion: let us suppose that Fodor were to demonstrate to everyone’s satisfaction that he could account for the mental-semantic properties of some mental states in terms of MR-semantic properties of primitives in Mentalese, where those MR-semantic properties were defined in terms of Fodor’s causal covariation account. On Fodor’s own account, however, those primitives make up only a small portion of possible expressions in Mentalese. To get to the level of the Mentalese representation of ‘The gloves and umbrella are on the bureau,’ one must combine semantic atoms according to syntactically-based rules. But, assuming that ‘*A*’ and ‘*B*’ are semantic atoms in Mentalese, by virtue of what does a Mentalese string such as [*A*-&-*B*] come to mean ‘*A* and *B*’?

If we were talking about discursive symbols, one would say that the syntactically-determined schema [___&....] has a certain semantic *function* in composing the meanings of larger expressions out of the meanings of their constituents. Now even if there is a notion of ‘syntax’ that can be understood apart from conventions—a notion that is just a matter of rules for legal concatenations (though ‘legality’ certainly *seems* like a conventional notion)—that kind of ‘syntax’ will not buy the kind of syntactically-based semantic function needed. Rules for concatenation, however complex, just do not do anything to determine what semantic function such concatenation performs. (This, I think, is one of the differences between the syntactic descriptions a descriptive linguist would be inclined to recognize as legitimate, such as ‘mass noun’ or ‘subordinating conjunction’ from ‘syntax’ that amounts to nothing more than concatenation: real syntactic categories can and often are characterized by *function* within a language game.)

In the case of the syntactic properties of discursive symbols, the general answer of how complexes get their meaning is fairly simple: the compositional rules (i.e., the rules for the semantic function of syntactic schemata) are conventional in nature, a matter of shared understandings and shared practices among players in the same language-game. But of course this option is simply not available if the ‘language’ is Mentalese. However the complexes get their meanings, it must not depend on conventions. For conventions require a community of agents privy to the conventions, and that would involve the theory in intentional homuncularism of a high order.

So while the appeal of building the generative and creative features of language into Mentalese is understandable, it is not clear now it could be accomplished. In natural languages, it is achieved by means of conventions that endow syntactic schemata with semantic functions. If CTM is to achieve the same result, it must find some other, non-conventional, way to endow equivalence classes of concatenations with semantic functions. The question of *how*, or even *if* this might be accomplished seems quite baffling. Perhaps even more baffling is the fact that, to the best of my knowledge, it has never been addressed in print by any of CTM’s advocates.[10] I do not rule out the possibility that there might be a way of closing this gap—it is always hard to argue that something *could not* be done. But it does strike me that here, even more than in the case of semantic properties of lexical primitives, it seems unlikely that one could come up with a non-conventional way of doing what is accomplished in natural language by means of conventions.

12. Science Without Sufficient Conditions

It would seem, then, that CTM has succeeded in neither of its philosophical tasks. It has not produced an account of intentionality in representational terms, but merely a promissory note for such an account. Nor can it make good on its claim to vindicate intentional psychology until it does so. Should we conclude, then, that computationalism is bankrupt as a paradigm for a psychology of cognition? It seems to me that the answer to this depends upon what we take psychology to be in the business of doing. If we take it that the job of psychology is to provide what we might call a ‘strong naturalization’ of the mental by way of providing naturalistic conditions that are metaphysically sufficient for mental states, then it seems very likely that computationalism cannot provide a viable paradigm for psychology. But neither is it clear that any other paradigm could do this either. In particular, it seems dubious that one could provide metaphysically sufficient conditions for consciousness, qualia or the phenomenological side of thought in naturalistic terms. But why should we expect psychology to do so? As I have argued in Horst (1992, 1996), we might expect that psychology should be committed only to a much weaker project: that of (a) specifying the form of the relations between mental states in formally precise terms, and (b) specifying the brain mechanisms through which mental states and processes are realized, where ‘realization’ is a relation that carries no metaphysical overtones. In brief, it seems to matter a great deal for purposes of a scientific psychology that there be formally exact models of the mental and that there be *some* sort of systematic mind/body relations; but it matters very little what the precise metaphysical nature of the relationship might be—it surely need not be such as to explain the essence of the mental. And, plausibly, computational psychology provides a format for doing each of these things: the technical resources of computer science provide resources for describing psychological processes in functional terms. And specifying an algorithm that has the right formal properties to support a cognitive process can provide important clues to the localization of such functions in the brain. Computational psychology is thus an interesting research format for psychology regardless of whether it delivers any philosophical goods. Whether it turns out in the end to be an *apt* model for the mind, and whether it competes successfully with other frameworks, like the nonlinear dynamic models of connectionists, is another question. My point here is merely that my objections to the philosophical claims of CTM do not themselves compromise more modest applications of the computer paradigm in empirical research and theory.

13. Conclusion

CTM is an attempt to solve certain problems about the mind by postulating a language of thought. The move holds substantial appeal. It appears to reduce two problems (mental meaning and symbolic meaning) to one (symbolic meaning). It attempts to extend familiar resources from the Chomskian revolution in linguistics to thought. And it appears to provide a way of vindicating psychological explanation in the intentional idiom. These appearances of progress, however, are illusory. The basic difficulty lies in the fact that the notions of *symbol*, *syntax* and *semantics* are paradigm-driven, and the paradigm instances are all convention- and intention-laden to the core. Any hopes for a successful explanation of the semantics of mental states, on the other hand, lies in providing sufficient conditions for (mental-)semantic properties in a fashion that is not thus dependent upon notions of convention or intention. Because the very *notions* of symbol, syntax and (semiotic-)semantics are convention-laden, it simply will not do to assume that the issue is merely one of finding a way of *endowing* mental representations with semantic properties in a non-conventional way. Rather, the issue is one of articulating what kind of properties these are supposed to be, and showing how they could account for mental-semantics. Likewise, it is not enough to say that semantic composition takes place in the mind, as in language, via syntax. For the syntactically-based rules in natural language for semantic composition are essentially dependent upon conventions that link syntactic schemata with semantic functions. It is not at all clear that one could achieve the same end through non-conventional means, nor even that one could meaningfully speak of arrangements of representations as ‘syntactic’ while avoiding the conventionality of syntactic categories of linguistic signs. The most plausible way of adapting the representational/computational approach to cognitive science is to abandon the close connections with semiotic notions and adopt a modest top-down strategy that treats the semantic properties of mental states as data and seeks to find the structures through which they are realized without any commitment to explaining them thereby. This yields no account of intentionality, but does seem to provide a framework for cognitivist research.

Notes

[1] Parts of this article were conceived during the NEH Summer Seminar on Mental Representation conducted at the University of Arizona, Tucson, in the summer of 1991. Special thanks go to Robert Cummins, the director of that Summer Seminar, who read both an earlier draft of this paper and a much longer manuscript dealing with the same subjects. Thanks also to Richard DeWitt of Fairfield University, who read an earlier draft, and to Kenneth Sayre of Notre Dame, who played a substantial role in my refinement of many of the ideas presented herein.

[2] Grice's notion of 'natural meaning' may by now be 'familiar' to philosophers, but I think my analysis later in the paper will have effect of revealing that it is simply *homonymous* with the other uses of the word 'meaning'.

[3] I have recently discovered a very similar articulation of a notion of conceptual dependence or priority in Stephen Schiffer's *Meaning*. The definition as stated is not completely perspicuous. Conceptual dependence is a transitive relation: it may be that X is analyzed in terms of Y and Y in terms of Z. In such a case, X is conceptually dependent upon Z even if Z would not occur in the most natural *articulation* of an analysis of X. That is, sometimes, for purposes of analysis of X, it might make best sense to mention Y in the analysis rather than Z, even though Y and hence X are both conceptually dependent upon Z.

[4] In Horst (1990, 1996) it is also suggested that there is a usage of 'symbol', particularly with regards to formal systems, in which it is syntactic typing that is crucial. The word 'counter' is there employed to replace this usage of 'symbol.' Pierce employs words such as 'semanteme' and 'syntagneme' to make the same distinction.

[5] This notion of interpretability-in-principle is more difficult to develop for markers than for signifiers. (See below.) The notion is discussed more thoroughly in Horst (1996), and hence is given little justification here.

[6] It is important to realize that this is an idealization. Change the voltage coming from your wall socket *significantly* and your computer will behave differently. Its behavior will seem like gibberish to you, but it is exhibiting a different functional architecture. The digital description of the machine treats things like voltage level as constant, and hence is an idealization, the way gravitational laws abstract away from influence of mechanical force and electromagnetism.

[7] Fodor (1975, p. 78) does mention the possibility of homonymy, at least with respect to the word 'representation', but finds it 'hard to care much how this question should be answered.'

[8] Several additional problems are discussed in Horst (1990, 1996).

[9] Cummins (1989) makes a similar distinction between 'the problem of meaning' and 'the problem of meaningfulness.'

[10] Or perhaps not so baffling. I, myself, was unaware of this problem for CTM until Rob Cummins pointed it out to me at an NEH Summer Seminar in 1991. I am thus quite indebted for this idea to Professor Cummins, who read a larger manuscript of mine on symbols, intentionality and CTM and pointed out to me that my claims about the conventionality of syntax bore upon this aspect of CTM, which I had previously overlooked, and which he had for some time felt needed to be addressed.

References

Block, N. (1986), 'Advertisement for a Semantics for Psychology', *Midwest Studies in Philosophy* 10, pp. 615–678.

Cummins, R. (1983), *The Nature of Psychological Explanation*, Cambridge, MA: MIT Press/Bradford Books.

Cummins R. (1989), *Meaning and Mental Representation*, Cambridge, MA: MIT Press/Bradford Books.

Dennett, D. (1987), 'Evolution, Error, and Intentionality', in D. Dennett, ed., *The Intentional Stance*, Cambridge, MA: MIT Press.

Dretske, F. (1981), *Knowledge and the Flow of Information*, Cambridge, MA: MIT Press/Bradford Books.

Dretske, F. (1986), 'Misrepresentation', in R. Bodgan, ed., *Belief*, Oxford University Press.

Dretske, F. (1988), *Explaining Behavior*, Cambridge, MA: MIT Press/Bradford Books.

Field, H. (1978), 'Mental Representation', *Erkentniss* 13, pp. 9–61.

Fodor, J. (1975), *The Language of Thought*, New York: Thomas Crowell.

Fodor, J. (1980), 'Methodological Solipsism Considered as a Research Strategy in Cognitive Science', *Behavioral and Brain Sciences* 3 (1980), pp. 63–73.

Fodor, J. (1981), *Representations*, Cambridge, MA: Bradford Books/MIT Press.

Fodor, J. (1987), *Psychosemantics*, Cambridge, MA: Bradford Books.

Fodor, J. (1990), *A Theory of Content and Other Essays*, Cambridge, MA: Bradford Books, MIT Press.

Fodor, J. (1994), *The Elm and the Expert*, Cambridge, MA: Bradford Books, MIT Press.

Haugeland, J, ed. (1981), *Mind Design*, Cambridge, MA: MIT Press/Bradford Books.

Horst, S. (1990), Symbols, Computation and Intentionality: A Critique of the Computational Theory of Mind. Doctoral Dissertation, University of Notre Dame.

Horst, S. (1992), 'Notions of 'Representation' and the Diverging Interests of Philosophy and Empirical Science', Proceedings of the 1992 Conference on Representation at SUNY Buffalo.

Horst, S. (1996), *Symbols, Computation and Intentionality: A Critique of the Computational Theory of Mind*. Los Angeles: University of California Press.

Loar, B. 'Conceptual Role and Truth Conditions', *Notre Dame Journal of Formal Logic* 23, pp. 272–283.

Millikan, R. (1984), *Language, Thought and other Biological Categories*, Cambridge, MA: MIT Press.

Millikan, R. (1986), 'Thoughts Without Laws: Cognitive Science Without Content', *Philosophical Review* 95, pp. 47–80.

Papineau, D. (1985), 'Representation and Explanation', *Philosophy of Science* 51, pp. 550–572.

Putnam, H. (1960), 'Minds and Machines', in S. Hook, ed., *Dimensions of Mind*, New York: New York University Press.

Putnam, S. (1961), 'Brains and Behavior', originally read as part of the program of the American Association for the Advancement of Science, Section L, December, 1961, printed in N. Block, ed., *Readings in the Philosophy of Psychology*, Cambridge, MA: Harvard University Press, 1980.

Pylyshyn, Z. (1980), 'Computation and Cognition: Issues in the Foundation of Cognitive Science', *The Behavioral and Brain Sciences* 3, pp. 111–32.

Sayre, K. (1986), 'Intentionality and Information Processing' *Behavioral and Brain Sciences* 9, No. 1 (1986).

Sayre, K. (1987), 'Cognitive Science and the Problem of Semantic Content', *Synthese* 70, pp. 247–269.

Searle, J. (1980), 'Minds, Brains and Programs', *Behavioral and Brain Sciences* 3, pp. 417–424.

Searle, J. (1983), *Intentionality*, Cambridge, England: Cambridge University Press.

Searle, J. (1984), *Minds, Brains and Science*, Cambridge, MA: Harvard University Press.

Schiffer, S. (1972), *Meaning*, Oxford, England: Clarendon Press.

Stich, S. (1983), *From Folk Psychology to Cognitive Science*, Cambridge, MA: MIT Press.